DOCUMENT RESUME

ED 137 358 95 TM 006 159

AUTHOR Rezmovic, Victor
TITLE The Secondary Analysis of the Cali Project.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
PUB DATE [Apr 77]
CONTRACT NIE-C-74-0015
NOTE 16p.; Paper presented at the Annual Meeting of the American Educational Research Association (61st, New York, New York, April 4-8, 1977)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Data Analysis; *Evaluation Methods; Measurement Techniques; *Program Evaluation
IDENTIFIERS Project Cali; *Secondary Analysis

ABSTRACT

The Project on Secondary Analysis at Northwestern University is funded to (1) test and develop new methods of evaluating educational programs, and (2) reanalyze existing evaluation data to assure that estimates of program effects are as unbiased and unequivocal as possible. This paper examines the topic of secondary analysis and describes some of the strategies that have been applied to Project Cali data. Project Cali is an intervention project designed to evaluate the impact of increased nutrition, remedial education, and medical care of malnourished children. Aspects of this data base which make it ideal for secondary analysis are described. The results of several of these techniques are presented in terms of their significance in interpreting the Cali program and in terms of their importance with respect to methodological issues related to evaluation research. (RC)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED137358

AERA 6.09

The Secondary Analysis of the Cali Project[1]

Victor Rezmovic
Department of Psychology
Northwestern University
Evanston, Illinois 60201

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Paper presented at the Annual Meeting of The American Educational Research Association, April, 1977, New York, New York.

TM006 159

The work which I will be describing is one of the research activities being conducted as part of the Project on Secondary Analysis (PSA) at Northwestern University. The PSA is being funded by a grant from the National Institute of Education and is "designed to (a) test and develop new methods of evaluating educational programs, and (b) reanalyze existing evaluation data to assure that estimates of program effects are as unbiased and unequivocal as possible" (Boruch, Wortman, and DeGracie, 1975). To date, reanalyses have been performed on such programs as Headstart (Magidson, 1977; Rindskopf and Wolins, 1977), Title I (Rindskopf and Wolins, 1977), ESAP (Alsip, 1977), Middlestart (Boruch, Magidson, Davis, 1975) and the Alum Rock Voucher Experiment (Wortman, Reichardt, and St. Pierre, 1977).

With few exceptions, not much systematic effort has gone into examining the topic of secondary analysis. Although there are some good examples of secondary analysis efforts in social science research such as the Mosteller and Moynihan (1972) volume examining the Coleman data, and Cook's (1975) book on the reanalysis of Sesame Street, the unique features of each secondary evaluation make them difficult to classify.

An excellent general exposition of the topic of secondary analysis and secondary evaluation is presented by Thomas Cook (1974). Cook's paper presents a 2x2x2 classification of secondary analysis models. The eight cells in Cook's table are generated by considering whether an analysis involves (1) a single data set or multiple data sets, (2) data reanalysis or no data reanalysis, (3) after-the-fact secondary analysis or simultaneous secondary analysis. Cook provides examples for each of his eight cells but also makes explicit that not all secondary analyses will fit into only one cell.

The approach by Rezmovic and Rezmovic (1976) examines one cell in Cook's (1974) matrix, the case involving reanalysis of a single data set after the fact. This is the situation which has been typically encountered in the PSA. Rezmovic and Rezmovic also present a 2x2x2 approach to secondary analysis considering whether the secondary analyst (1) used new statistical methods or repeated the same methods as the original investigator, (2) was concerned with the same variables as the primary analyst or examined different variables and (3) whether the reanalysis addressed the same issue as the original evaluation or a set of different hypothesis. As with Cook's matrix, the Rezmovic and Rezmovic matrix may place a secondary analysis project into several cells. The reanalysis of the Cali Project, for example, encompassed many of the cells in the matrix.

A recent paper by Hendricks and Wortman (1977), although presenting no taxonomy, looks at secondary analysis as a decision-making process in which the secondary analyst makes key choices different from those of the original investigator. The main point of this paper is that decisions about analysis strategies should not be seen as better or worse when compared to one another, but that different approaches can answer similar questions of interest. People do secondary analysis for a variety of reasons. Housewives, for example, routinely check their grocery bills after shopping at the supermarket to make sure they have not been cheated. University professors delight in examining statistical reports and finding a numerical inconsistency in a data table provided by another colleague, or one of their graduate students. Researchers reexamine studies conducted by others hoping to find evidence for a theoretical position to which they subscribe. The motives and methods of those conducting a reanalysis are varied. I would like to mention some of the more common reasons.

The first step in performing secondary analysis frequently involves redoing the analyses of the original investigator. This usually means using raw unaggregated data to reproduce means, variances, correlations, and cross tabulations. This is done to insure that the data sent from primary investigator to the secondary analyst are the same. It is not unusual to hear of data tapes prepared at one computer installation producing slightly different data when output at a second center. Another reason for reproducing these basic analyses is to make sure that these measures were correctly computed. The quality of results reported in journal articles and published reports is not well known. In one study attempting to measure data quality, Wolins (1962) discovered conceptual and computational errors in three of seven studies which he reanalyzed. Discrepancies between the original results and the reanalysis may be due to sources such as keypunching errors, varying the number of subjects used to compute the statistics, or mislabeling certain variables.

Research is sometimes conducted under time pressure and with limited resources in computers, programs, personnel, and finances. Given these realistic constraints, an evaluation may not be totally adequate. The secondary analyst, in such an instance, may want to perform different analyses which answer the research questions in more depth. One advantage of being a secondary analyst is not having to perform under the strict time constraints that primary researchers are frequently subjected to.

Using statistics to evaluate the success or failure of a program sometimes involves arbitrary decisions about which particular statistical test to use. Different statistical methods often provide essentially the same information, e.g., using analysis of variance instead of regression.

However, in some cases statistics can be used inappropriately, resulting in different results than if the more "correct" method had been used. The most common instance of this is the use of a statistic when one or more of the assumptions underlying the statistic have been violated. The classic paper by Campbell and Erlebacher (1970) demonstrates how regression effects may have made the original Headstart evaluation look harmful. A more recent demonstration of this type of problem is provided by Rindskopf and Wolins (1977), showing how inappropriate use of the analysis of covariance may have biased the results of an evaluation of the Title I education programs.

In certain instances, the motivation behind a secondary analysis may be to clarify certain issues which may be vague in the original report. For example, when statistical results are not significant at the .05 level, the secondary analyst may wish to use a statistical test which is more powerful to insure that the lack of significance was not due to a test low in power. For example, the analysis of covariance increases the power of an F test over that produced by the regular anlaysis of variance model. Another example might involve reanalyzing a project in which correlations were reported but direction of causality is unclear. Using a structural equation model might provide more information about the direction of cause. Other instances of this type of anlaysis might involve testing for interactions not examined by the original investigator.

There have been several examples of secondary evaluations which have attempted to refute the results presented by the primary analyst. The most notable contemporary example is the research on the genetic components of intelligence. The analyses by Kamin (1974) have attempted to refute conclusions drawn by Jensen (1969). Another example is the Elashoff and

Snow (1971) reanalysis of Rosenthal and Jacobson (1968) data on teacher expectancy.

Many data sets contain information on variables which are not analyzed, or even reported by the primary investigator. For example, a data set may contain information on program participants which relates SES to income but these variables may not interest the original investigator. A secondary analyst interested in the relation between income and SES would want this data set but his analysis would not involve secondary analysis per se. One possibility for primary investigators who have collected large amounts of variables which remain unanalyzed is to advertise their data sets so others may use the information which is present. Creation of data archives would particularly facilitate this type of effort.

Many new methods for evaluating programs exist. These methods are derived by statisticians and mathematicians and rarely have immediate relevance for evaluation researchers. Testing out these methods with existing data sets serves a valuable purpose in that members of the research community are informed as to how to use these techniques and what the advantages of the techniques may be. In many cases new methods are used with simulated data and these have the advantage of letting the researcher know the true model underlying the data. However, when using empirically generated data certain problems may arise which would not arise using the simulated data.

Reanalysis has the advantage of teaching students and researchers about how to do research. Analytic methods need to be practiced to be understood. Availability of data banks of project information is a valuable tool in the teaching process. Currently, many courses in research methods

use a data base and have students' assignments keyed to analyses which are performed on this data base.

When performing secondary analyses on multiple data sets one may be able to resolve contradictory findings that plague a research area. For example, an education program implemented at two sites may bring success at one and failure at the other. A secondary analysis might determine what factors differ between the sites and how these factors may have prevented success at the second site. Another example of this is the meta-analysis approach taken by Gene Glass (1976) in which results of studies in an area are aggregated to give a more coherent picture of the findings.

Integrating contradictory findings may serve a second function, i.e., allowing the secondary analyst to make statements about the policy relevance of certain research results. Secondary analysis may come too late to provide information for decision-makers at certain key points in the decision-making process. Accumulated findings, however, will certainly have an impact at later stages as policy is continually shifting with new administrations. Up to now, I've addressed general issues in performing a secondary analysis. I would now like to direct my comments to the secondary analysis of the Cali data set in particular.

Often times the most difficult part of conducting a secondary analysis is the act of getting the data. In the case of the Cali data, many of the typical problems which are encountered were not present. This is due, in part, to the long relationship between Northwestern University and the Cali researchers. Many members of the Northwestern community have served and continue to serve as advisors to the project. This ongoing relationship has kept a channel of communication open between Cali and Evanston

and when the PSA began it was natural that the Cali data would be a data set which would be reanalyzed.

Another reason for the smooth relationship between Cali and Northwestern University lies in the different interests of the two groups. Researchers in the Cali Project, for example, are interested in questions relating to the effectiveness of the program. Does the treatment improve cognitive performance and reduce malnutrition? Their interests center around substantive matters relating to the world of the undernourished Colombian child. Those of us involved in the secondary analysis are educational researchers whose main interests concern the development of new methods to measure program impact.

One might argue that our reanalysis is not a true "secondary analysis" in that we have dealt with new issues not related to the original analyses. Such criticism is partially justified because many of our analyses are really primary ones and much of our work might be considered concurrent evaluation as opposed to after-the-fact evaluation. In one sense, our analyses are extensions of the primary analysis because of the richness of the data base. No single research team could hope to use all the information collected by this project.

Based on our experiences with the Cali Project we would urge primary investigators, who collect evaluation data at great cost and effort, to share their data with others so that maximum benefits accrue to the research community.

The PSA has been seeking educational data sets amenable to secondary analysis. We feel that the Cali data provide us with good opportunities to perform various analyses. There are several characteristics of the

data base which we find useful in our secondary analysis efforts.

For example, the longitudinal nature of the data, i.e., repeated measurements over a five-year period have allowed us to use within subjects analyses which have more statistical power. In addition, the repeated measurement have enabled us to study differing growth profiles in the experimental groups as a function of treatment. Since the Cali study employed a true experimental design with random assignments of subjects to treatment conditions, it has maximized both the primary and secondary analysts' abilities for drawing unequivocal conclusions about treatment effects. In addition, its inclusion of a non-equivalent control group enabled us to compare the randomized controls with the non-equivalent controls. This comparison is important in helping to establish the direction and magnitude of bias produced when non-equivalent controls are utilized, the typical case in most social experiments.

Multiple measures of achievement, SES, and medical status are valuable because they facilitate the use of many sophisticated techniques which require multiple measurements of a single latent variable. These multiple measures also allow us to measure the effects of the program in more specific terms and to outline the relative success of the treatment in various contexts of the child's modes of performance.

We must applaud the members of the project for the care taken to insure that the quality and documentation of data were adequate for our purposes. Although we received over 130 data cards for each of over 300 cases, there has been no difficulty locating the information of interest. The data system instituted in Cali has been very successful and much effort has been devoted to minimizing both transcription and keypunching errors.

Our initial efforts of the secondary analysis entailed replicating the original means, variances, and correlations produced by the original Cali researchers. We were able to do this successfully and were quite certain that our results matched theirs. Although we did discover very minor errors in keypunching, such as values which exceeded the maximum possible score, these errors were so infrequent and affected the results so slightly, that we were confident about the results of the primary report.

We are happy to say that all the methods that we used for measuring program effect led to conclusions which were similar to those of Dr. McKay and his colleagues.

Our analyses which included repeated measures ANOVA, univariate and multivariate ANOVA, and discriminant analysis, all supported the conclusion that the program works. These strong results allowed us to concentrate on questions of methodological interest rather than trying to establish how well the program worked.

Many sophisticated methods of data anlaysis exist which can be used to assess the effects of treatment programs. Frequently, these methods are first described by mathematical statisticians and are not immediately available to educational researchers. This is either because the computer software to perform these analyses is not available or because these techniques have not been reliably demonstrated in educational contexts. One of the aims of the PSA has been to demonstrate the applications of new methods of analysis to conventional educational research situations. I would like to mention several of these.

Our analyses looked at the Cali data of height and weight as repeated measurements over a five-year period using seven observations during this time.

In dealing with the issue of repeated measures analysis, two approaches were contrasted. The univariate analysis of repeated measures (or mixed model, as it is also known) looks at whether profiles for the various treatment groups are parallel. If so, we conclude that there was a treatment effect. In the multivariate approach, we treat the repeated observations as a set of dependent variables and the difference between the groups can be tested. The univariate approach makes the assumption that the pooled within-group Σ, exhibits compound symmetry, i.e., that the same construct is being measured at each time point with equal reliability. The multivariate approach makes no assumptions about Σ. The difference between the two approaches is that when the assumptions of the univariate approach hold it yields a more powerful analysis. In the multivariate approach, many degrees of freedom must be used at estimating the within-group covariance matrix. In our reanalysis, we performed both types of analysis and examined how the univariate statistics could be extracted from the multivariate results.

Before the Rasch model approach was taken, an aim of the original Cali analysis was to combine several cognitive measurements into a general ability factor and to compare groups across the time periods on this factor. This approach, while included in an earlier report, was later rejected because it did not consider the question of factorial invariance between groups and across occasions. Recently Jöreskog (1971) has described a technique for testing factorial invariance. Using Jöreskog's technique, we focused on the question of factorial invariance among treatment groups at the last treatment period. Analyses were performed which examined the factorial structure of the WISC-R subtests used in five randomly assigned Cali groups.

Results of this analysis yielded a stable factor pattern with two factors (verbal and performance) invariant among all groups. Using the Jöreskog approach, we were also able to compute group factor means.

Two problems which have plagued researchers concern the effect of measurement error on statistical results and (2) making causal inferences in correlational research. There has been a recent interest in the area of structural equation models which deal with these two issues. In particular, we were interested in examining the causal relationship between nutritional status and intelligence. A causal model was hypothesized in which we could examine the effects of nutritional status on intelligence. The next presentation will examine some of these analyses in more detail.

Our reanalysis of the Cali data attempted to use new statistical techniques to answer questions of educational significance, and to provide the research community with examples of new methods which can be used in evaluation contexts. We hope our efforts will encourage other educational researchers to further these goals.

Footnotes

1. Work on this manuscript was supported by Contract No. C-74-0115 from the National Institute of Education.

References

Alsip, J. E. A review of the Emergency School Assistance Program and its evaluation. Research report, Division of Methodology and Evaluation Research, Northwestern University, 1977.

Boruch, R. F., Davis, S., and Magidson, J. Interim report: Secondary analysis of Middlestart. Evaluation Research Report NIE-12. Division of Methodology and Evaluation Research, Northwestern University, 1975.

Boruch, R. F., Wortman, P. M., and DeGracie, J. Executive Summary: Secondary Analysis of Educational Program Evaluations. Presented at the annual meeting, American Educational Research Association, April, 1975.

Cook, T. D. The potential and limitations of secondary evaluations. In M. W. Apple, M. J. Subkoviak, and H. S. Lufle (Eds.) Educational Evaluation: Analysis and Responsibility. Berkeley, California: McCutchon, 1974.

Cook, T. D., Appleton, H., Conner, R. F., Shaffer, A., Tamkin, G., and Weber, S. J. Sesame Street Revisited. New York: Russell Sage Foundation, 1975.

Elashoff, J. D. and Snow, R. E. Pygmalion Reconsidered. Worthington, Ohio: Charles A. Jones, 1971.

Glass, A. Primary, secondary, and meta-analysis of research. Educational Researcher, 1976, 5(10), 3-8.

Hendricks, M. and Wortman, P. M. Justifying secondary analysis: Arbitrary components in large-scale data analysis. Paper to be presented at the Annual Meeting of the American Psychological Association, 1977.

Jensen, A. R. How much can we boost I.Q. and scholastic achievement? Harvard Educational Review, 39, (1969), 1-123.

Jöreskog, K. G. Simultaneous factor analysis in several populations. Psychometrika, 1971, 36, 409-426.

Kamin, L. J. The science and politics of I.Q. New York: Wiley, 1974.

Magidson, J. Towards a causal model approach for adjusting for pre-existing differences in the nonequivalent control group situation: A general alternative to ANCOVA. Research Report, Division of Methodology and Evaluation Research, 1977.

Mosteller, F. and Moynihan, D. P. (Eds.) On equality of educational opportunity. New York: Random House, 1972.

Rezmovic, V. and Rezmovic, E. Components of Secondary Analysis. Research Bulletin, Division of Methodology and Evaluation Research, Northwestern University, 1976.

Rindskopf, D. and Wolins, L. Using regression correlation methods to adjust for pretreatment differences. Evaluation Research Report NIE-01RW. Division of Methodology and Evaluation Research, Northwestern University, 1977.

Rosenthal, R. and Jacobson, L. Pygmalion in the classroom: Teacher expectation and pupils' intellectual development. New York: Holt, Rinehart, and Winston, 1968.

Wolins, L. Responsibility for raw data. American Psychologist, 1962, 17, 657-658.

Wortman, P. M., Reichardt, C. S., and St. Pierre, R. G. The First Year of the Educational Voucher Demonstration: A secondary analysis of student achievement test scores. Education and Urban Society, 1977, (in press).